EDITORIAL

NUCLEIC ACIDS RESEARCH ANNUAL WEB SERVER ISSUE IN 2009

The 2009 Web Server Issue of *Nucleic Acids Research* is the seventh in a series of annual special issues dedicated to web-based software resources for analysis and visualization of molecular biology data. It is freely available online under *NAR*'s open access policy. The present issue reports on 112 web servers.

TOPICS

This year's special emphasis was on metagenomics, molecular network and pathway analysis and biological text mining. Fourteen papers deal with these topics. Another 14 papers involve web services, biocomputing workbenches or bioinformatics tools for such tasks as clustering, genome browsing, deep sequencing and homology search. By far, the largest number of papers covers DNA and RNA (17) and proteins (36). The remainder cover a variety of topics, including gene annotation, gene set enrichment analysis, phylogeny, microarrays and SNPs.

Also included in the present issue is the Bioinformatics Links Directory 2009 update by Michelle Brazas, Francis Ouellette and their colleagues at the Ontario Institute for Cancer Research. The directory, at http://bioinformatics.ca/links_directory, is a searchable compilation of web servers published in this and previous Web Server issues together with other useful tools, databases and resources for life sciences research.

INSTRUCTIONS FOR SUBMISSIONS

To streamline the review process, authors are required to send a one-page summary of their web server to the editor, Dr Gary Benson (narwbsrv@bu.edu), for pre-approval prior to manuscript submission. For the 2009 issue, 282 summaries were submitted and 141, exactly 50%, were approved for manuscript submission. Of those approved, 112, or nearly 80%, were accepted for publication.

Review of a summary includes evaluation of the proposal and extensive testing of web server functionality. The key criteria for pre-approval are high scientific quality, wide interest, the ability to do computations on user-submitted data and a well-designed, well-implemented and fully functional web site. Note that there is a minimum 2-year interval before publication in the Web Server issue for web servers, or essentially similar web servers, that have been the subject of a previous publication, including publication in journals other than *NAR*.

With respect to the web site, the following are guidelines for approval. It should have an easy-to-find submission page with a simple mechanism for loading test data and setting test parameters. The preferred method is one-click loading using Javascript or a similar mechanism. Also acceptable, but less preferred, is data available through a link next to the data submission box. This requirement simplifies the review process for the editor and the referees and provides potential users with a quick way to examine and judge a web server's features. Additional mechanisms that assist the user in submitting data should be implemented where appropriate. If the user can submit data that could be downloaded programmatically from a source website, for example a pdb structure file or a GenBank sequence file, then the web server should provide automatic download of that data once the user has entered the appropriate identifier.

Output of the web server should be dynamic and rich in detail. Wherever possible, links should be provided to supporting evidence used in calculations and/or external databases containing additional information. Numerical, textual and visual output should be mixed and any visualization tools that add information or increase the user's understanding should be utilized (e.g. the Java plug-in tools jmol for structure visualization and jalview for sequence alignment visualization). Note that a web server with output that consists merely of a few numerical values, a static spreadsheet or a compressed file will not be approved (although download options for static files should be an option).

Many web servers provide time-consuming analysis and are not able to return results immediately. In that case, a mechanism for returning results to the user should be implemented. Although notification by email is straightforward and might be provided as an option, many users balk at revealing their identity through email. The preferred method is to return a web link to the results, at the time of data submission, which the user can then copy and access at a later time. This link should ideally report the status of the job (queued, running or finished). Even if the user provides an email address for results, they should be provided on a webpage (in the dynamic form mentioned above) rather than mailed as large files that might be rejected as spam by email programs.

The web site should be supported by an extensive help section or tutorial that guides the user through the submission process, contains details about input file formats and parameters, and importantly, explains the meaning of the output. Whenever possible, the help pages should link to dynamic output examples similar to those provided by the web site.

Any proposal for a web server that is *predictive* must include details on validation of predictions from new data not used in training. N-fold cross-validation methods will not be considered sufficient. Details should include size and

composition of the validation dataset (number of positive and negative cases), and several measures of predictive performance, including sensitivity, specificity and precision. Proposals are regularly rejected for lack of adequate prediction validation information.

Many summaries are rejected because the web sites are clearly not designed to accept user-submitted data. This applies to those established primarily for lookup or exploration in a dataset, or serve the function of 'data integrators', even if the data are not stored locally. Authors of web sites that provide novel data should consider the *NAR* Database Issue as a possible venue (see the instructions at http://www.oxfordjournals.org/our_journals/nar/for_authors/msprep_database.html).

Proposals that describe a novel analysis method are generally not appropriate for the Web Server issue because limited space makes adequate method description and validation problematic. Authors of such methods might instead consider sending their manuscript to *NAR* as a regular computational biology paper (see the instructions for authors at http://www.oxfordjournals.org/our_journals/nar/for_authors/criteria_scope.html#Computational%20Biology).

NEW FOR 2010

Stand-alone programs for high-throughput data

Very high volume experimental data are becoming more common, for example, with the advent of next generation sequencing. In response to this, the 2010 Web Server issue will inaugurate a new section for stand-alone (non-web server) programs that analyze such data. While web servers are ideal for ease-of-use, especially for new or inexperienced users, high volume data present two significant problems: (i) excessive time is required to upload the data (and limited web-browser upload capacity may make such uploads impossible), and (ii) processing at a centralized (and often academic) computing resource may overtax that resource to the point where it becomes unable to serve the target audience. In contrast, stand-alone programs can be run locally, thus distributing the computing load and simplifying data upload.

In order to maintain the ease-of-use standards expected of web servers published in *NAR*, a stand-alone program must meet the following criteria: (i) it must deal primarily with high-throughput data, defined as input data that commonly equal or exceed 10 MB, (ii) it must run on at least two of the three common operating systems (linux, mac and pc), (iii) it must be self-installing, that is, it must have a well tested, 'one click' type installation and (iv) it must have a well-designed graphical user interface similar to that of a web browser. At this time, exclusively command line programs will not be considered.

Web services

The 2010 issue will also inaugurate a separate section for short papers describing popular web services, that is, the automated analysis that can be utilized programmatically rather than through manual interaction with a web server.

Special emphasis

For the 2010 issue, analysis of next generation sequencing data will be added to the continuing special focus on metagenomics, network and pathway analysis and biological text mining.

DEADLINES FOR 2010

Authors wishing to submit manuscripts for the 2010 Web Server issue must submit their one-page proposal along with the URL address of the fully functional website to narwbsrv@bu.edu by 31 December 2009. Detailed instructions and requirements are presented at http://www.oxfordjournals.org/nar/for_authors/submission_webserver.html. This information should be consulted before sending in the summary. The deadline for submission of articles is 31 January 2010.